A Weight-Selection Strategy on Training Deep Neural Networks for Imbalanced Classification

7/4/2017

Antonio Sze-To, Andrew K.C. Wong

Center for Pattern Analysis and Machine Intelligence (CPAMI)

Department of Systems Design Engineering

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

Introduction

1. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (2015)
2. 2. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al.: Mastering the game of go with deep neural networks and tree search. Nature 529(7587), 484–489 (2016)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- **Deep Learning** [1] has achieved **superhuman** performance in many places such as beating the world champion in **GO** [2].

- Beats Lee Sedol 9p in 4:1
- Beats Ke Jie 9p in 3:0

Google's AlphaGo beats the world's best Go player, again

Chinese Go player Ke Jie plays a match against Google's artificial intelligence program, AlphaGo, during the Future of Go Summit in Wuzhen in eastern China's Zhejiang Province, Thursday, May 25, 2017. (Chinatopix via AP)

Introduction

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- Not only GO, **Deep Learning** has **achieved superhuman performance** in a variety of tasks

19th Feb, 2015

Microsoft's Deep Learning Project Outperforms Humans In Image Recognition

Michael Thomsen, CONTRIBUTOR
I write about tech, video games, science and culture. **FULL BIO**
Opinions expressed by Forbes Contributors are their own.

24th Feb, 2016

Intelligent Machines

Google Unveils Neural Network with "Superhuman" Ability to Determine the Location of Almost Any Image

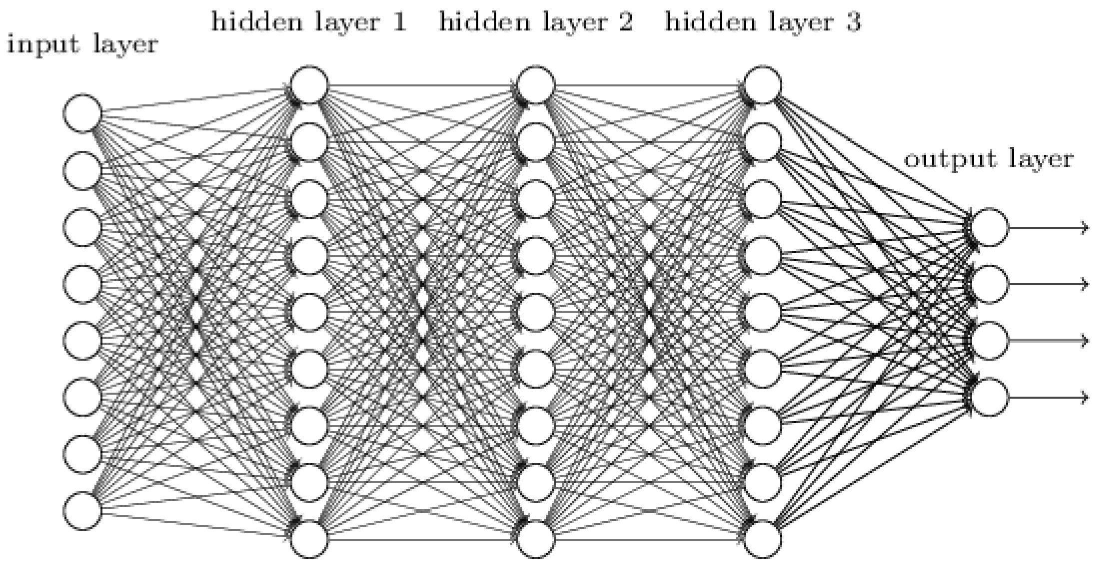
25th Sept, 2015

Artificial Intelligence Can Beat Humans at 31 Atari Games

Introduction

3. Bengio, Y., et al.: Learning deep architectures for AI. Foundations and trends. Mach. Learn. 2(1), 1–127 (2009)

- **Deep Neural Networks (DNN)**, defined as artificial neuron networks with at least 3 hidden layers [3]).



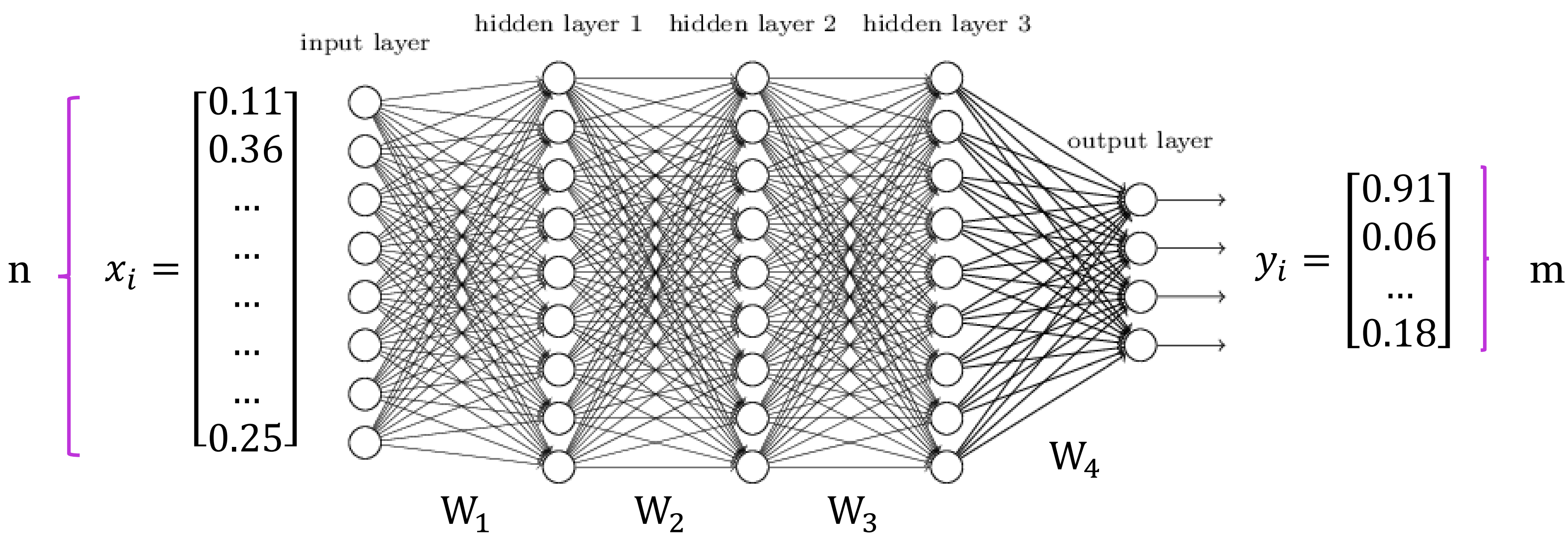
Source: http://neuralnetworksanddeeplearning.com/chap5.html

- **Deep Learning** refers to the use of **Deep Neural Networks (DNN)** in learning tasks.

How DNN works?

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- Given a n-dimensional vector x_i, a DNN outputs a m-directional vector y_i, with the weights being the model parameters



How DNN works?

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- Given a n-dimensional vector x_i, a DNN outputs a m-directional vector y_i, with the weights being the model parameters

$$\text{n} \left\{ \quad x_i = \begin{bmatrix} 0.11 \\ 0.36 \\ \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ 0.25 \end{bmatrix} \right.$$

The DNN can be abstracted as

$$g_\theta(x_i) = y_i$$

$$\theta = \{W_1, W_2, \dots W_4\}$$

$$\left. y_i = \begin{bmatrix} 0.91 \\ 0.06 \\ \dots \\ 0.18 \end{bmatrix} \right\} \quad \text{m}$$

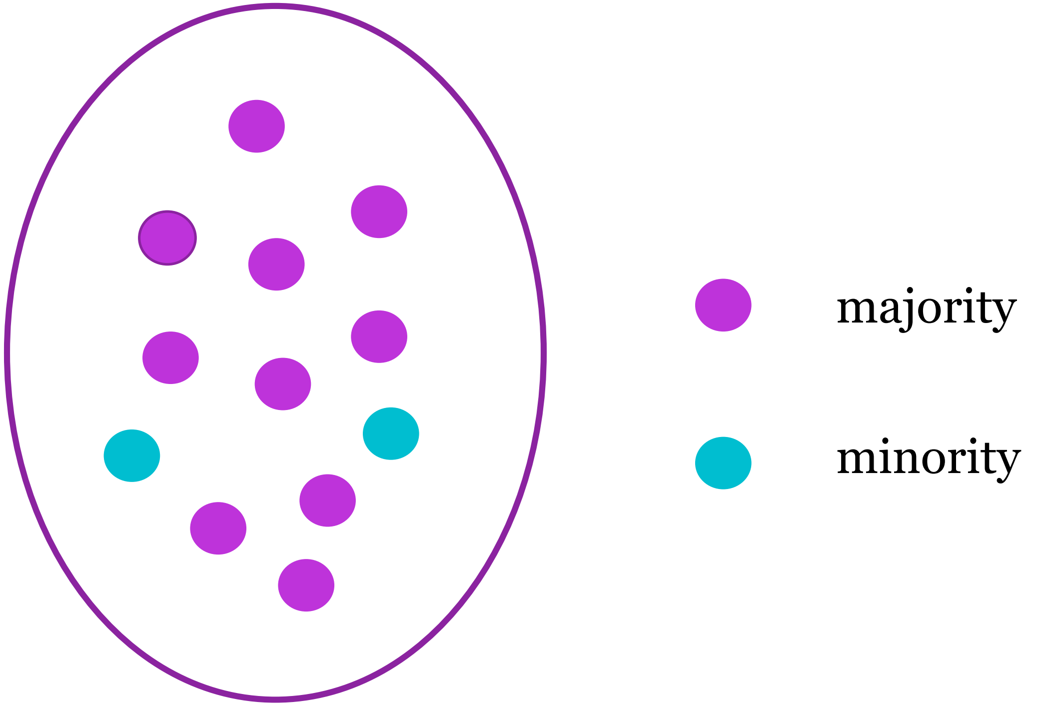
UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Introduction

5. Sun, Y., Wong, A.K., Kamel, M.S.: Classification of imbalanced data: a review. Int. J. Pattern Recogn. Artif. Intell. 23(04), 687–719 (2009)
6. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21(9), 1263–1284 (2009)
7. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: review of methods and applications. Expert Syst. Appl. 73, 220–239 (2017)

- Imbalanced classification [5–7]

refers to the problem that one class contains a much smaller number of samples than the other classes in classification.



Introduction

6. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21(9), 1263–1284 (2009)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- Such phenomenon exists in many real-world machine learning applications [6], e.g.
    - fraud detection,
    - medical diagnostics
    - equipment failure prediction

Related work

9. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357 (2002)
10. Ting, K.M.: A comparative study of cost-sensitive boosting algorithms. In: In Proceedings of the 17th International Conference on Machine Learning. Citeseer (2000)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- Throughout the years, many methods have been developed for tacking imbalanced classification.
- They can be categorized into 2 types:

(1) resampling [9],

- which changes the balance between classes

(2) cost-sensitive learning [10]

- which assigns higher misclassification costs to the minority class

Related work

14. Shen, W., Wang, X., Wang, Y., Bai, X., Zhang, Z.: Deepcontour: a deep convolutional feature learned by positive-sharing loss for contour detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3982–3991 (2015)
15. Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., Kennedy, P.J.: Training deep neural networks on imbalanced data sets. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 4368–4374. IEEE (2016)
16. Ng, W.W., Zeng, G., Zhang, J., Yeung, D.S., Pedrycz, W.: Dual autoencoders features for imbalance classification problem. Pattern Recogn. 60, 875–889 (2016)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- However, studies on training DNN on imbalanced classification are still scanty.

- Recent trend includes
  - regularizing the loss function [14,15] during training
  - representation learning [16]

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Motivation and Objective

- In this study, we explore a new strategy to train DNN for imbalanced classification based on weight selection.

- To our knowledge, it is the 1st time to examine a weight-selection strategy on training DNN on imbalanced classification.

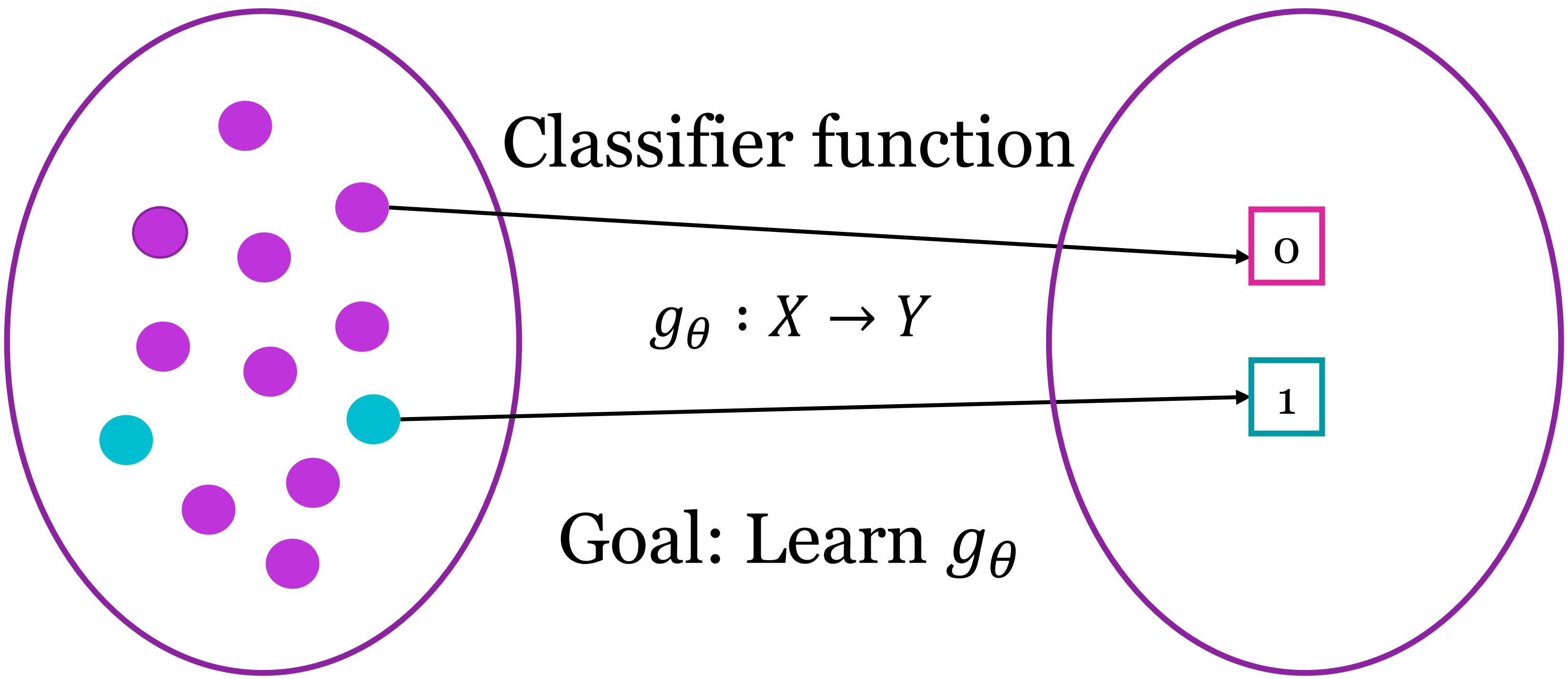
- Our objective is to investigate if such strategy can improve the predictive performance of DNN on imbalanced classification, even when the DNN are trained on a reduced training set.

Problem Definition: Supervised Learning

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

Training samples

Class labels



$X = R^n$ (n-dimensional space)

$Y = \{0, 1\}$

Problem Definition: Supervised Learning

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

Given $S_{train} = \{(x_i, y_i)\}_{i=1}^{N}$, the goal is to learn a function $g_\theta: X \rightarrow Y$ such that

$$E_{(x,y)\sim D}[L(g_\theta(x_i); y_i)] \qquad (1)$$

is minimized, where

- $L(g_\theta(x_i); y_i)$ is a loss (error) function
  - It measures the loss (error) between $g_\theta(x_i)$ and y_i.
- $x_i \in X = R^n$, $y_i \in Y = \{0,1, \ldots C\}$
- D is the distribution over $X \times Y$

Problem Definition: Supervised Learning

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

Given $S_{train} = \{(x_i, y_i)\}_{i=1}^{N}$, the goal is to learn a function $g_\theta : X \to Y$ such that

$$E_{(x,y)\sim D}[L(g_\theta(x_i); y_i)] \qquad (1)$$

In practice, Equation (1) would be approximated over a set of testing samples $S_{test} = \{(x_i, y_i)\}_{i=1}^{M}$ as Equation (2)

$$Test_{g_\theta}(S_{test}) = \frac{1}{M}\sum_{i=1}^{M}[L(g_\theta(x_i); y_i)] \qquad (2)$$

Training DNN (Standard Strategy)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- By applying **back-propagation** algorithm on S_{train}, a set of weights can be obtained, i.e. a new θ, to achieve a smaller loss value in each iteration

Algorithm 1. DNN

Input: a set of training samples S_{train}, a function g, initialized model parameter θ_1

Output: a model parameter θ corresponds to function g.
initialize $\theta = \theta_1$
for t = 1 to epochs **do**
  $\theta_{t+1} = \text{backprop}(S_{train}, \theta_t)$
  $\theta = \theta_{t+1}$
end for
return θ

For 1 to epochs do:
  Train on S_{train} to update weights
End For

Training DNN (Validation-Loss Strategy)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- Here we propose a new strategy known as **Validation-Loss (VL) strategy**.

- It is to split the input training set S_{train} into 2 sets,
  - T_{train}: train the DNN to update weights,
  - $T_{validate}$: validate the performance

- The weights that render the minimum loss value on the validation set would be selected

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Training DNN (Validation-Loss Strategy)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

Algorithm 2 DNN+VL

Input: a set of training samples S_{train}, a function g, initialized model parameter θ_1
Output: a model parameter θ corresponds to function g.
split S_{train} into T_{train} and $T_{validate}$
initialize $\theta = \theta_1$
initialize $v_{loss} = Test_{g_\theta}(T_{validate})$
for t = 1 to epochs **do**
  $\theta_{t+1} = \text{backprop}(T_{train}, \theta_t)$
  if $v_{loss} > Test_{g_\theta}(T_{validate})$ **then**
    $v_{loss} = Test_{g_\theta}(T_{validate})$
    $\theta = \theta_{t+1}$
  end if
end for
return θ

Split S_{train} to T_{train} and $T_{validate}$
For 1 to epochs do:
  Train on S_{train} to update weights
  Test on $T_{validate}$
  Save the weights if improved
End For

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Datasets

18. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE 86(11), 2278–2324 (1998)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- MNIST dataset [18] is a handwritten digit dataset, containing 60,000 training images and 10,000 testing images.

- Each image is associated with a distinct digit as its class. There are 10 digit classes (0, 1, 2,. . . , 9) in total.

- All the images were resized to 28x28=784 pixels

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Datasets

19. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE 86(11), 2278–2324 (1998)

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

Table 1. A summary of the 10 imbalanced image datasets obtained from MNIST [19]. Each dataset has two classes: -ve and +ve classes, where the +ve class represents a digit (minority class) and the -ve class represents the remaining digits (majority class). The imbalance ratios (ImR) were calculated via dividing the no. of -ve samples by the no. of +ve samples. The higher the ImR is, the more imbalance the dataset is.

	Training -ve	Training +ve	Training ImR	Testing -ve	Testing +ve	Testing ImR
Dataset-0	54077	5923	9.13	9020	980	9.20
Dataset-1	53258	6742	7.90	8865	1135	7.81
Dataset-2	54042	5958	9.07	8968	1032	8.69
Dataset-3	53869	6131	8.79	8990	1010	8.90
Dataset-4	54158	5842	9.27	9018	982	9.18
Dataset-5	54759	5421	10.10	9108	982	10.21
Dataset-6	54082	5918	9.14	9042	958	9.44
Dataset-7	53735	6265	8.58	8972	1028	8.73
Dataset-8	54149	5851	9.25	9026	974	9.27
Dataset-9	50000	5949	9.09	8991	1009	8.91

Experimental Setup

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

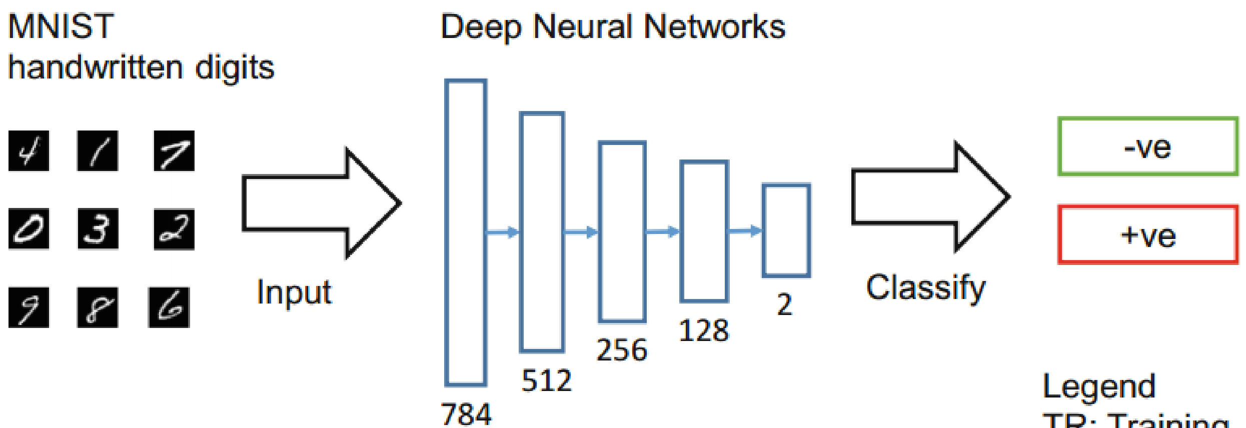
Classifier	Description
1	SVM with a RBF kernel
2	Decision tree (CART)
3	DNN: a DNN trained by a **standard strategy** on the entire training set of each dataset (which contains 60,000 images)
4	DNN-CL: a DNN trained by a **standard strategy** with **cost-sensitive learning**, i.e. setting the misclassification cost of the minority class to the majority class as ImR (Imbalanced Ratio) to 1;
5	DNN-VL: a DNN trained by the **Validation-Loss** strategy training only on a randomly selected 48,000 (80%) images from the training set of each dataset for weight updates, and the remaining 12,000 (20%) images for weight selection.

These classifiers were tested on the testing 10,000 images of each dataset.

Experimental Setup

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

a) Imbalance classification using Deep Neural Networks



Legend
TR: Training
VA: Validation
TE: Testing

b) Standard Strategy

TR | TE

For 1 to epochs do:
    Train on the TR set to update weights
End
Test on the TE set

c) Proposed Strategy

TR | VA | TE

For 1 to epochs do:
    Train on the TR set to update weights
    Validate the performance on the VE set
    Save the weights if the VE's results improve
End
Load the saved weights
Test on the TE set

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Experimental Results

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

Table 2. A comparison of the prediction performance among 5 classifiers in terms of the area under the receiver operating characteristic (ROC) curve in **testing**. Two-tailed Wilcoxon Signed-Rank Test was conducted on the ROC values obtained by DNN and DNN-VL, as well as the ROC values obtained by DNN-CL and DNN-VL. In both tests, we obtained a p-value of $\mathbf{0.00512} < \mathbf{0.05}$.

Testing ROC	SVM (RBF)	Decision tree	DNN	DNN-CL	DNN-VL
Dataset-0	0.99864	0.96504	0.99549	0.99555	**0.99917**
Dataset-1	0.99857	0.97613	0.99671	0.99436	**0.99921**
Dataset-2	0.98468	0.91820	0.99297	0.99088	**0.99874**
Dataset-3	0.98835	0.91336	0.99299	0.99558	**0.99865**
Dataset-4	0.99323	0.92614	0.99383	0.99363	**0.99817**
Dataset-5	0.98109	0.92016	0.98830	0.99558	**0.99921**
Dataset-6	0.99577	0.95732	0.99120	0.99261	**0.99880**
Dataset-7	0.98999	0.94502	0.99244	0.98820	**0.99602**
Dataset-8	0.98351	0.89631	0.98670	0.97505	**0.99835**
Dataset-9	0.97748	0.90388	0.98679	0.98825	**0.99575**

Experimental Results

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

Table 3. A comparison of the prediction performance among 5 classifiers in terms of the area the under precision-recall curve (PRC) in **testing**. Two-tailed Wilcoxon Signed-Rank Test was conducted on the PRC values obtained by DNN and DNN-VL, as well as the PRC values obtained by DNN-CL and DNN-VL. In both tests, we obtained a p-value of $\mathbf{0.00512} < \mathbf{0.05}$.

Testing PRC	SVM (RBF)	Decision tree	DNN	DNN-CL	DNN-VL
Dataset-0	0.99630	0.96852	0.99709	0.99526	**0.99847**
Dataset-1	0.99715	0.97564	0.99749	0.99591	**0.99851**
Dataset-2	0.97197	0.92987	0.99457	0.99034	**0.99696**
Dataset-3	0.98099	0.91459	0.99400	0.99251	**0.99708**
Dataset-4	0.98123	0.93479	0.99355	0.99020	**0.99675**
Dataset-5	0.97172	0.92722	0.99148	0.99414	**0.99741**
Dataset-6	0.99136	0.96317	0.99349	0.99367	**0.99740**
Dataset-7	0.98191	0.94941	0.99232	0.98696	**0.99429**
Dataset-8	0.96266	0.90656	0.99249	0.97733	**0.99520**
Dataset-9	0.95822	0.91355	0.99095	0.98462	**0.99363**

Conclusion

CPAMI
INTELLIGENT RESEARCH EXCELLENCE

- In this study, we proposed a new strategy, known as weight-selection strategy, for training Deep Neural Networks (DNN) on imbalanced classification.
- To our knowledge, it is the 1st systematic study to examine a weight-selection strategy on training DNN on imbalanced classification.
- Illustrated by experiments on 10 imbalanced datasets obtained from MNIST, the DNN trained on a reduced training set by the new strategy outperformed the DNN trained by a standard strategy and cost-sensitive learning with statistical significance ($p=0.00512$)
- Future work includes incorporating the proposed strategy with existing techniques (e.g. resampling) to further improve the performance.

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Availability

- All experimental results and source codes are available in https://github.com/antoniosehk/WSDeepNN

content	Add files via upload	3 months ago
.gitignore	Initial commit	4 months ago
DataUtils.py	1st commit	4 months ago
LICENSE	Initial commit	4 months ago
MLModel.py	Double-checked results	10 days ago
README.md	Update README.md	3 months ago
WSDeepNN.py	Double-checked results	10 days ago
experimetal_results.xlsx	Double-checked results	10 days ago

README.md

WSDeepNN

A Weight-Selection Strategy for training Deep Neural Networks for Imbalanced Classification

Motivation

Deep Neural Networks (DNN) have recently received great attention due to their superior performance in many machining-learning problems. However, the use of DNN is still impeded, if the input data is imbalanced. Imbalanced classification refers to the problem that one class contains a much smaller number of samples than the others in classification. It poses a great challenge to existing classifiers including DNN, due to the difficulty in recognizing the minority class. So far, there are still